

RESEARCH ARTICLE

Evaluating the robustness of snow climate indicators using a unique set of parallel snow measurement series

Moritz Buchmann^{1,2}  | Michael Begert³  | Stefan Brönnimann^{2,4} | Christoph Marty¹

¹Snow and Permafrost, WSL-Institute for Snow- and Avalanche Research SLF, Davos, Switzerland

²Oeschger Centre for Climate Change Research, University of Bern, Switzerland

³Climate Division, Federal Office of Meteorology and Climatology MeteoSwiss, Zürich, Switzerland

⁴Institute of Geography, University of Bern, Bern, Switzerland

Correspondence

Moritz Buchmann, SLF Davos, Flüelastrasse 11, CH-7260 Davos, Switzerland,
Email: moritz.buchmann@slf.ch

Funding information

Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: SNF-200021L_175920

Abstract

Snow on the ground is an important climate variable which is normally measured either as snow depth or height of new snow. Like any other meteorological variable, manually measured snow is prone to local influences, changes in the environment or procedure of the measurements. In order to investigate the robustness of snow measurement series towards such non-climatic changes, a unique set of parallel manual snow measurements over 25 years from 23 station pairs between 490 and 1800 m a.s.l. was compiled. A sensitivity analysis based on typical snow climate indicators (e.g., mean snow depth, sum of new snow) from these parallel time series was carried out to find the most robust snow climate indicators for climatological analyses. Results show that there are only small differences in the sensitivity of the various snow climate indicators with regards to local changes. However, the indicators number of days with snow on the ground as well as the maximum snow depth are least affected by local influences and changes at station level. Median values of all station pairs reveal relative differences of about 7% for the number of days with snow cover and 11–16% for all other indicators. However, in extreme cases, the deviations within a single station pair can reach 25–40%.

KEYWORDS

climate, climate indicator, parallel time series, snow, snow measurements

1 | INTRODUCTION

Snow has multiple implications for a wide range of areas like ecology, economy and society: ranging from plants, animals, habitats and cycles of life (Jonas *et al.*, 2008; Wipf *et al.*, 2009; Resano-Mayor *et al.*, 2019) to winter tourism, hydro power, fresh water availability, floods, avalanches and climate feedbacks (Marty, 2008; Scherrer *et al.*, 2012; Marcolini *et al.*, 2017b; Schmucki *et al.*, 2017). Modern climatological studies of past snow trends in the Alps have a

relatively short history in science with first studies starting in the 1990s (Beniston *et al.*, 1994; Spreitzhofer, 1999). Since then, various studies have focused on a variety of aspects (such as trends, variations, or forecasts and modelling) in all the above-mentioned fields, although there are still questions regarding the quality and the representativeness of the actual measurements and corresponding climatological time series.

The longer a time series, the more likely it has experienced breaks due to changes of observer, location,

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2020 The Authors International Journal of Climatology published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society.

instrument or procedure. Such breaks can affect the quality of the data as mentioned by Aguilar *et al.* (2003) and Della-Marta and Wanner (2006). This is a fundamental issue in climate sciences which has been well-studied for temperature and precipitation, (e.g., Begert *et al.*, 2005; Scherrer *et al.*, 2013; Acquafredda *et al.*, 2019; Guenzi *et al.*, 2020) but very rarely addressed for snow (Marcolini *et al.*, 2017a). It is widely accepted that “snow” measured in either height of new snow (HN) or snow depth (HS) can be spatially quite heterogeneous and the measurements are prone to local influences (e.g., Neumann *et al.*, 2006; López-Moreno *et al.*, 2015). Snow is not only a function of temperature and precipitation but also dependent on elevation, exposure to wind, solar irradiation, and to a large extent on its instability on a micro-structural level (e.g., settlement, metamorphism). To address such issues, sensitivity analyses of parallel time series are paramount and have been conducted for temperature and precipitation (Acquafredda *et al.*, 2016; Gubler *et al.*, 2017; Hunziker *et al.*, 2017). For snow only few studies exist (e.g., Acquafredda *et al.*, 2015; Baronetti *et al.*, 2019) analysing parallel snow series. However, data from two different measurement techniques were used (manual and automatic), large distances (up to 20 km) were allowed within their station pairs and no impact on indicator series was investigated.

The present study benefits from the fact that in Switzerland snow is monitored by two independent institutions (WSL Institute for Snow and Avalanche Research Davos SLF and Federal Office of Meteorology and Climatology MeteoSwiss) and that the measurement principle has not changed since the beginning. This circumstance allows the exploitation of a carefully constructed unique set of long-term, daily, manual data of independent, parallel snow measurements. The aim is first to investigate the sensitivity of indicators, derived from snow depth and new snow with regards to local changes (in either location or observer). Throughout this paper, the influence of such changes (environment, instructions, observer) is not accounted for because of lack of trustful metadata. Nevertheless, this is exactly the point of this study – to find the least sensitive indicators regardless of any changes that might have occurred during the analysed time period. This last point has practical implications, as normally neither observer metadata nor environmental changes of a station itself are completely documented.

The results of this analysis are used to assess which snow indicators are most robust for climate studies, that is, least sensitive to local changes. This has practical implications for any further homogenisation approaches as well as for climate services where recommendations about trend analyses of usually un-homogenised snow time series can be improved. This will be accomplished by introducing and analysing derived snow climate

indicators, like mean snow depth, number of days with snowfall, etc. and assessing the stability by making use of the parallel long-term measurements.

This paper is organised as follows: Section 2 introduces the data set and outlines the statistical methods used for the analyses. Results are presented in Section 3, followed by a discussion in Section 4. Conclusions are drawn in Section 5.

2 | DATA AND METHODS

Switzerland consists of three major parts: Swiss Plateau in the north, Jura mountains in the west, and Prealps and Alps in the south. Topography and climate are consequently complex and diverse. Typical mean maximum annual snow depths range from 15 cm for the Swiss Plateau to 300 cm in the Alps, and 70–100 cm for inner Alpine valleys and pre-Alpine regions. Starting at an altitude of 1,200–1,500 m above sea level, precipitation during winter predominantly falls as snow, such that the area is often covered by a solid layer of snow for weeks, and even months at higher altitudes. Snowfall is relatively rare in the low-lying areas of western Switzerland (greater Geneva area) and northern Switzerland (greater Basel area) as well as in the lowland in the southern tip of Switzerland.

Daily operational, manual snow measurements usually entail at least two variables: height of new snow as a 24 h sum (HN) and snow depth (HS). HN has to be measured, as the difference in HS (retrieved, say in a 60 min interval) is usually not the amount of new snow accumulated in said interval due to settlement of the snow pack. In Switzerland, snow is measured since the late 19th century. Manual measurements, which are solely used in this study, are still conducted today, using basically the same instruments as in the beginning. Both MeteoSwiss and SLF maintain a network of manual snow observations, be it for slightly different objectives. For MeteoSwiss, as the Federal Office of Meteorology and Climatology, snow is one of many variables they are interested in and basically just one form of precipitation, in contrast to the SLF where snow and snowfall are important due to its main brief of avalanche forecasting. The station distribution of the two networks reflects that focus, as the MeteoSwiss-stations are located throughout Switzerland, whereas SLF-stations are distributed solely over the mountainous and alpine areas.

2.1 | Measurement procedures

Daily measurements are conducted each morning by reading off the value from a stake with centimetre scale (HS) and by taking three measurements with a ruler on

the snow measurement board (HN) both at 06:00 UTC (Haberkorn, 2019). Instructions vary slightly when it comes to reporting as SLF stations normally only measure between November and April (unless there is already or still snow present) in contrast to MeteoSwiss stations which measure all year round. Unlike MeteoSwiss, where HN smaller than 0.5 cm is recorded as zero, SLF allows to put in traces (fixed default value of 0.3 cm) for HN smaller than 0.5 cm. These traces have been subsequently set to zero in the data analysis in order to be consistent with MeteoSwiss procedures. Values are stored in two databases and have already been processed independently in terms of initial quality control and gap filling. But both data sets have been analysed again with a systematic manual quality control looking for gaps and implausibilities (see Section 2.2) prior to being used for the analyses in this study. However, according to MeteoSwiss and SLF experts, there is no publication available documenting the various QC methods.

2.2 | Parallel data set

Because of the various operational focuses, there are locations where both institutions operate (or have operated) an observer site within the same village at some period in time. To build the parallel station set, the candidates have to meet the following criteria: Roughly within the same village (± 3 km), similar elevation (± 100 m), data from November to April, independent data (as we cannot exclude instances in which data were copied), and parallel measurement series for at least 20 consecutive years. Stations can, but do not have to be part of the same network.

For simplification, we define the hydrological year as follows:

$$hyear_{1980} : = 1.10.1979 \text{ to } 30.9.1980 \quad (1)$$

and subsequently winter or snow season as the period from November to April.

Unfortunately, it is not just a question of comparing the coordinates from the two networks to build the station pairs. Stations can have two names and different coordinates but still be identical, due to the fact that the coordinates represent the meteorological station, rather than the actual location of the snow measurements, as they, for practical reasons often cannot be conducted too close to the measurement field of the automatic instruments. Additionally, in the past it was not deemed important to know the exact location of the snow measurements. Starting off with a rough list of possible station pairs, compiled on available metadata such as

current elevation and location, each case had to be analysed in detail in order to make sure the pair was independent. A possible pair of two existing stations has then to be checked for independent data and overlap.

Sometimes a station pair only appears to be independent, but contains in fact exactly the same data. This could happen for the following reasons: In order to fill in missing values for one station, sometimes the data from the corresponding “partner” station was simply copied during one or multiple short periods of time. Or one observer reports to both networks separately. Moreover, the search for station pairs proved to be cumbersome as approximately one third of the MeteoSwiss stations was not digitally available and had gaps that needed careful treatment first.

Station metadata such as coordinates and observer names could sometimes help solving the conundrum. Unfortunately, these data are not always available and trustworthy, especially for past changes. Even if available, observer names such as “Swiss Border Force” are still vague (e.g., station Santa Maria 1970s). Additionally, each institution has its own data base and therefore different or no quality codes that inhibits a simple query to check whether some data were copied in the first place.

To address these issues, a simple quality control mechanism is introduced. Stations are treated as independent if more than 60% of the data are not equal (without counting zeros) for any given winter (empirical value, gained by visual analysis). The above criteria yield a preliminary set of more than 55 possible station pairs. By selecting only station pairs where its members have more than 80% of data available for each annual winter season the number of possible station pairs is further reduced. Considering the length of the overlap and the benefit of looking at the same period, a dataset of 23 station pairs (see supplementary material Table S1) results for the 25-year period between 1980 to 2004 (see map in Figure 1).

Data is subjected to quality control focusing on implausibilities (like for example, snow depth decrease of 50 cm within one day) and consistencies between HN and HS. All cases of possible implausibilities and inconsistencies were manually checked. It is important to mention that the number of such cases were rare, which is not surprising as most of the time series have already been checked by the data owners. Missing values were interpolated by manually fitting the evolution derived from the best correlated neighbouring station using median ratios (again, the occurrences were rare, as the 80% cut-off meant that only high-quality stations [almost complete series] were selected in the first place). After the interpolation process all analysed time series were complete. However, not all inconsistencies can be

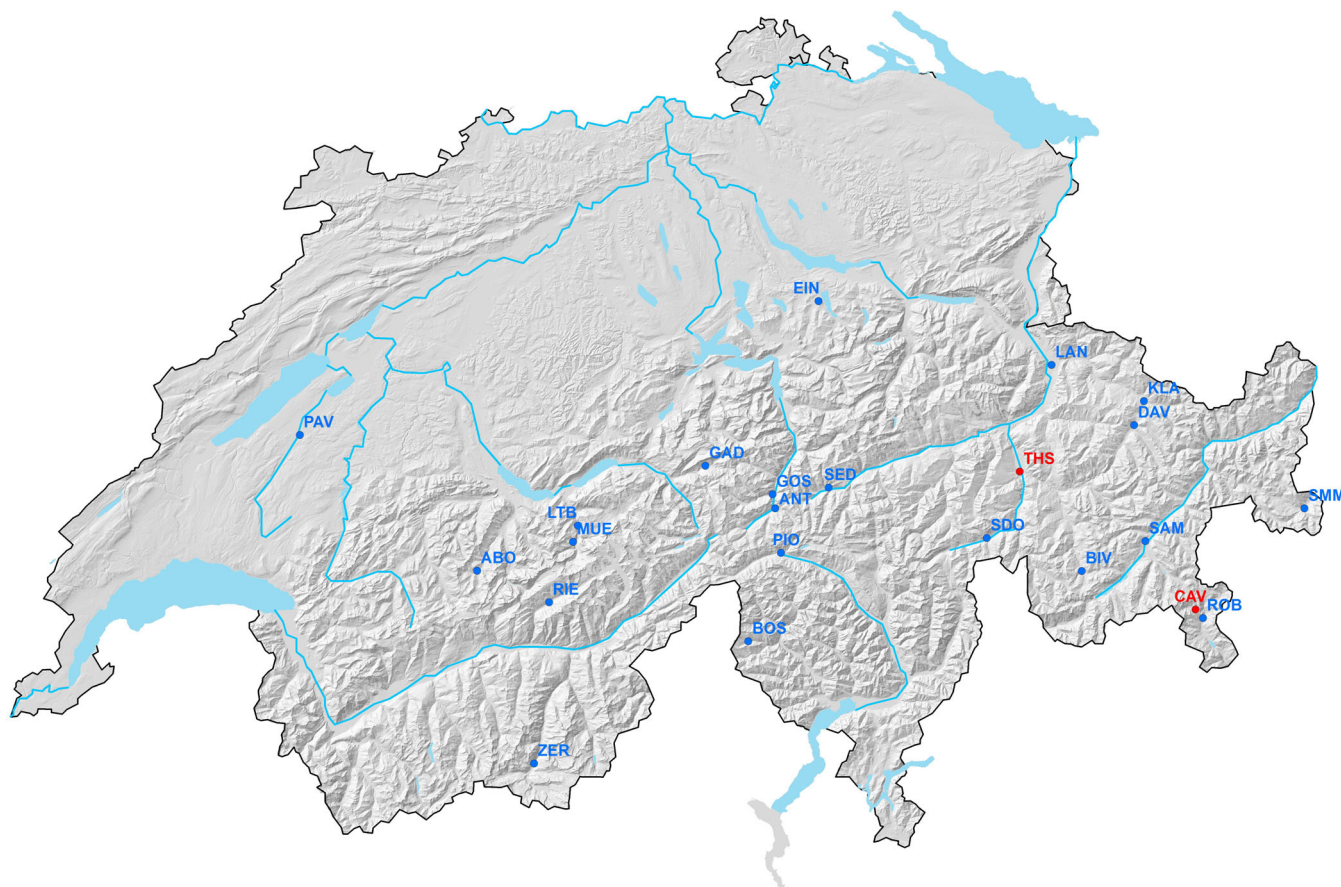


FIGURE 1 Map of Switzerland, showing the distribution of the station pairs used in this study. The red dots (THS and CAV) highlight station pairs that were excluded for the HN and trend analyses due to inadequate HN series [Colour figure can be viewed at wileyonlinelibrary.com]

addressed as MeteoSwiss stations assign HN to the previous, rather than the actual day. This issue is historically partly addressed in the observer form, in the digitisation and/or in quality control process by simply moving the HN time series by one day. For these reasons this shift is unfortunately not always constant for all stations, time periods and sometimes not even an entire winter season. All possible combinations of no shift, two-day shift or shifts in the wrong direction can occur. Sometimes, even HS series have been affected. However, these cases could be easily detected and have been corrected accordingly. Fortunately, the HN shift does not affect the calculations, as the indicators are defined as annual values. For two stations the HN series had to be omitted, because HN was most of the time just calculated from the difference between today's and yesterday's snow depth (see red stations in map in Figure 1), resulting in 23 station pairs for HS- and 21 for HN-indicators. Hereafter, only the term “23 station pairs” is used for improved readability. Finally, all station pair time series were visually checked for possible remaining issues.

2.3 | Snow climate indicators

To be able to compare the station pairs and carry out the sensitivity analysis, snow climate indicators are introduced and defined as annual values from the daily HS and HN measurements similar to WMO's Global Climate Indicators; only with focus on snow and calculated for each station pair and hydrological year. Only years that have valid values for both stations are used to determine the snow climate indicators, which are:

Number of days with HS of at least 1 cm (dHS1) is widely used to establish whether the ground is snow-covered, an important factor for ecologists and climate scientists alike. To account for the fact that dHS1 might be sensitive to the observation time and method, dHS5 with a 5 cm threshold is introduced, as $HS \geq 5$ cm can still be regularly observed at all stations in the data set (see Section 4.3). The same applies to number of days with HN of at least 1 or 5 cm (dHN1 and dHN5) which are mainly of interest for tourism, road maintenance and climatology. The maximum sum of three consecutive

FIGURE 2 Winter 1997 at station Adelboden: The bold red line shows HS and the black bars indicate days with HN. All visible bars below the 5 cm threshold (black horizontal line) are counted as dHN1, all that exceed the 5 cm threshold as dHN5 (analogue for dHS1 and dHS5). The red horizontal line depicts HSavg. HSmax, HNmax, HN3max, and HNsum (shown as cumulative sum) are also indicated [Colour figure can be viewed at wileyonlinelibrary.com]

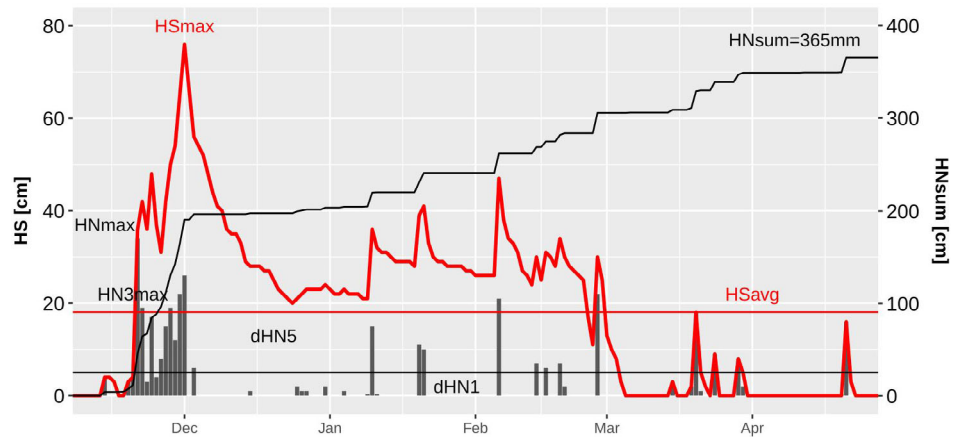


TABLE 1 Summary of the snow climate indicators

Indicator	Description	Unit
HSavg	Mean HS	cm
HSmax	Max HS	cm
HNmax	Max HN	cm
HNsum	Sum HN	cm
HN3max	Max sum over three days	cm
dHS1	Number of days with HS > 0 cm	days
dHS5	Number of days with HS > =5 cm	days
dHN1	Number of days with HN > 0 cm	days
dHN5	Number of days with HN > =5 cm	days

snowfall days (HN3max) is of particular interest for avalanche forecasts, maximum snow depth (HSmax) has implications for national snow load codes and civil engineering, average snow depth (HSavg) is of interest for climate sciences, whereas the sum of new snow (HNsum) is more suited for hydrological or climatological analyses (refer to Figure 2 and Table 1).

2.4 | Statistical methods

2.4.1 | Correlations and absolute differences

Correlations (COR) between station pairs for all climate indicators are calculated using Spearman's rank coefficient as this method is less susceptible to outliers than Pearson's rho. Deviations within a station pair are expressed as relative percentage differences (RPD) between SLF (slf in the equations) and MeteoSwiss (mch) stations, because none of the two networks can be called a proper reference series.

$$RPD = \frac{mch - slf}{\text{mean}(mch, slf)} \quad (2)$$

RPD yield one annual value per station pair and indicator. Boxplots are used to compare the distribution of the RPD for all station pairs, separated for each snow climate indicator by using the mean RPD over all 25 years for each station pair.

To give a sense for the absolute scale of deviations, root mean squared errors (RMSE) are also calculated; always mch minus slf, analogue to RPD.

$$RMSE = \sqrt{\text{mean}((mch - slf)^2)} \quad (3)$$

2.4.2 | Median absolute deviations

In order to quantify the variability of the correlations and relative percentage deviations across all station pairs, median absolute deviations (MAD) are favoured over coefficients of variation due to their robustness with regards to outliers.

$$MAD = \text{median}(|X_i - \text{median}(X)|); \text{with } X \in \{COR, RPD, \dots\} \quad (4)$$

2.4.3 | Relative changes

In order to assess the comparability of the long-term temporal evolution among station pairs changes are calculated using the Theil-Sen linear slope (Theil, 1950; Sen, 1968) and the non-parametric Mann-Kendall test (Mann, 1945; Kendall, 1975). Relative changes (RC) are defined as:

$$RC := \frac{\text{fitted value at the end} - \text{fitted value at the beginning}}{\text{median of the fitted values}} \quad (5)$$

To investigate the differences within a station pair, the differences in the (above) relative changes (DRC) were calculated as:

$$DRC := RC(mch) - RC(slf) \quad (6)$$

3 | RESULTS

3.1 | COR, RPD, and RMSE

The median correlation values for each of the nine indicators range from 0.81 for HNmax to 0.95 for HSavg (Figure 3). However, median values are higher than 0.85 for all snow climate indicators except HNmax. HNmax also reveals by far the highest variation among all station pairs. In contrast, HSavg and HSmax, show by far the

smallest variability. The outliers in Figure 3 consist of three different station pairs with no apparent connection.

Relative inter-pair deviations (Figure 4), expressed as median RPD for all station pairs range between 5% and 15%. The lowest values are clearly shown by dHS1 and dHS5, which together with HSmax also reveal small variation among the station pairs. On the other side, the largest variation result from dHN1 and dHN5. The outliers not shown in Figure 4 correspond to the lowest station pair in the data set. The visible outliers (HSmax, dHS1, and dHS5) relate to two snow-poor station pairs (PAV and ROB, see Table S1 for details).

In contrast to RPD, RMSE reveals the absolute deviation values (Figure 5). HNsum shows the largest RMSE with 24 cm and by far the largest variability. Median values for the other indicators range from 1.6 cm for HSmax to 3.5 cm for HN3max and from 2.2 days for dHN5 to 4.5 days for dHS5.

Because higher stations usually experience more snowfall, larger snow depths, and more snow days, possible elevation dependences were analysed for COR and RPD but no clear signal is found (see Figures S1–S4 and Table S2).

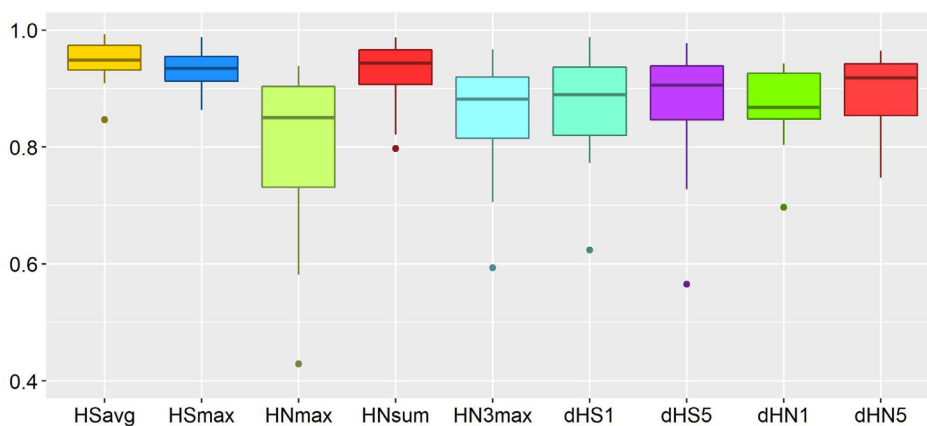


FIGURE 3 Spearman correlation for all station pairs grouped by snow climate indicators [Colour figure can be viewed at wileyonlinelibrary.com]

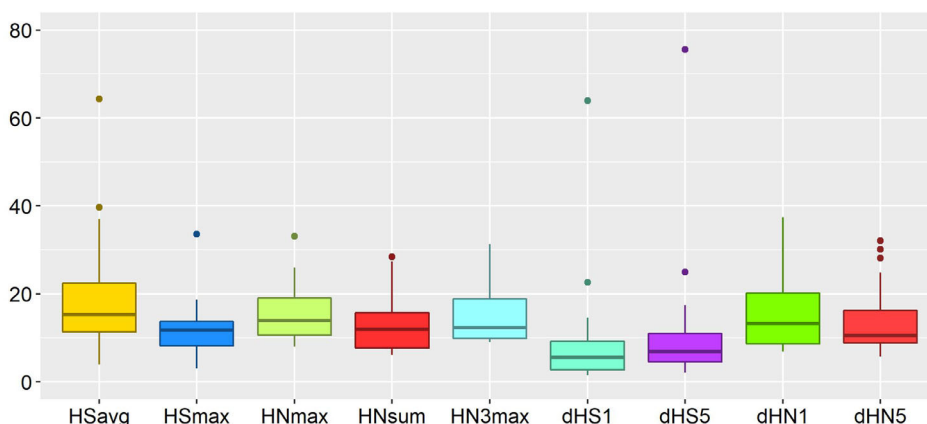


FIGURE 4 Comparison of mean RPD (in %) for each station pair and indicator. dHS1 and dHS5 both show median RPD smaller than 10% [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 5 RMSE for the various snow climate indicators. Note that HNsum values are an order of magnitude larger than the rest and so is its RMSE [Colour figure can be viewed at wileyonlinelibrary.com]

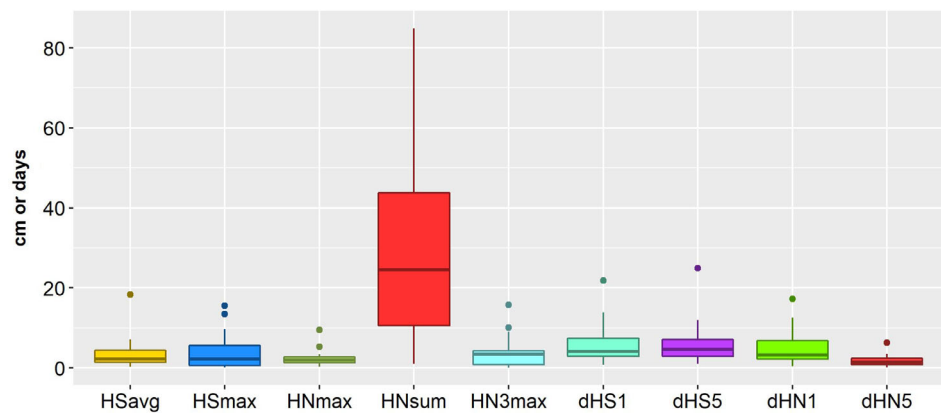


FIGURE 6 Relative changes (RC) for each station and indicator in %. Station pairs are ordered according to their mean HSavg from low (PAV) to high (ANT). Station names are defined in Table S1 [Colour figure can be viewed at wileyonlinelibrary.com]

3.2 | Relative changes

Figure 6 shows strength and direction of temporal RC for each station pair. Nearly all stations show negative RC for all indicators for the period under investigation. The sign of the changes agrees well within the station pairs; the only differences occur when one station has a value close to zero (see Figure 6 for details). The DRC (depicted in Figure 7) reveals median values for all station pairs below 10%. Furthermore, for a majority of stations and indicators, most values are below 20%, except for HNmax and dHN1. dHS1 and dHS5 show clearly the smallest values for the majority of stations, but several outliers which are associated with two snow-poor station pairs.

3.3 | MAD

Table 2 reveals the variability between the station pairs for COR, RPD, and DRC expressed as MAD in percentage values for all snow climate indicators. Larger values indicate a larger variability of the underlying metric and thus more variation across all station pairs (visible in Figures 3, 3, 4, and 7). MAD(COR) scores best for HSavg and HSmax with 4%. HNsum, dHS1, dHS5, dHN1, and dHN5 form the next group with values of 7%, followed by HNmax with 13%, illustrating the large spread in Figure 3. MAD(RPD) shows a more gradually distribution with values ranging from 4% to 11%. Again, HSmax, as well as dHS1 and dHS5 are associated with low values,

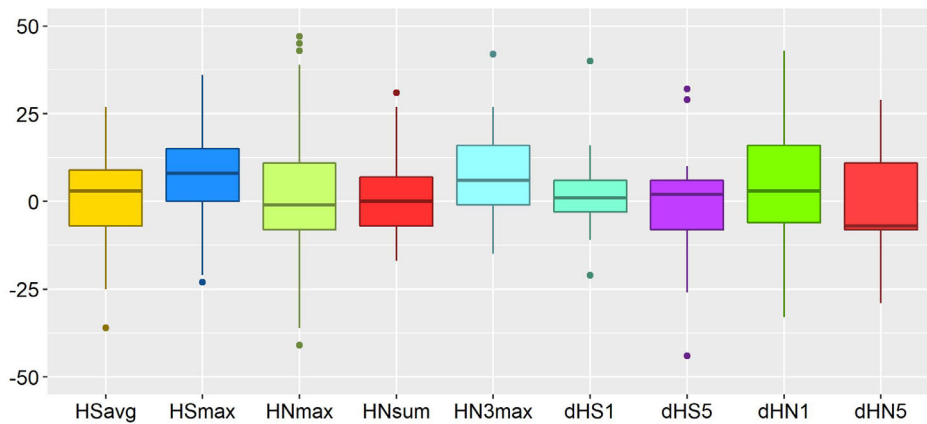


FIGURE 7 Differences in relative changes (DRC) for all snow climate indicators in % [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Comparison of the median absolute deviations for COR, RPD, and DRC for all snow climate indicators

Indicator	MAD COR [%]	MAD RPD [%]	MAD DRC [%]
HSavg	3	7	19
HSmax	3	4	12
HNmax	13	7	21
HNsum	7	8	12
HN3max	7	4	15
dHS1	7	4	9
dHS5	7	4	9
dHN1	7	8	18
dHN5	7	11	16

Note: Larger values indicate larger spreads (%) of the underlying metric. The two smallest values in each column are marked bold.

whereas here HNsum and dHN5 bring up the rear. MAD(DRC) values are generally higher than MAD(COR) and MAD(RPD) and like MAD(COR) separable into three groups: dHS1, dHS5 building the top with 9%, HSmax, HNsum, HN3max, and dHN5 in the middle with 12%–16% and dHN1, HSavg, and HNmax at the bottom with 19–21%.

4 | DISCUSSION

4.1 | Elevation dependences

The absence of any evident elevation dependences for COR, RPD, and DRC, determined with a simple linear regression (Figures S2–S4) and shown as coefficient of determination in Table S2, allows comparison and ranking of the snow climate indicators in order to find the most stable ones. However, due to the small sample size and unevenly distributed stations over the elevation

range of our data set, a thorough analysis is not feasible. Having the largest RPD associated with the lowest station pair with the smallest absolute values is of no surprise and explained by the calculation of RPD itself.

The mere existence of elevation dependences for the mean indicator values (see Figure S1) does not come unexpectedly as one would expect more snow at higher elevations due to lower temperatures. However, the contrast with respect to elevation among the various snow climate indicators is interesting. HN-indicators are first and foremost precipitation dependent. Above a certain altitude as mentioned by Morán-Tejeda *et al.* (2013), the temperature is generally low enough for the occurrence of snow. HS-indicators on the other hand also require precipitation, but are much more dependent on temperature.

4.2 | COR, RPD, and RMSE

The COR values are similar to the ones retrieved by Acquaforte *et al.* (2015) for the comparison of two independent precipitation networks and generally quite high (median COR > 0.8). The overall weaker correlations for HNmax can be explained by the fact that HNmax is one single event per year and not necessarily recorded at the same date within a station pair. The small-scale nature of precipitation itself, the short lifespan of weak snowfall events, as well as local influences (shade, sun, exact measurement time) can also have a bearing.

When looking at RPD, the counting variables cover both ends of the spectrum: dHS1 and dHS5 display the smallest values, whereas dHN1 and dHN5 show high values and the largest variability. This can be explained by the overall greater absolute number of days with snow cover compared to days with snowfall. Small precipitation events coupled with varying time of observation and exposure of the actual measurement site have a greater impact on dHN1 which is visible in the larger spread of

dHN1 compared to dHN5. Overall, the variability in dHN1 and dHN5 can be attributed to small-scale variations in precipitation. The indicators HSavg, HNmax, HNsum, and HN3max show very similar values and distribution patterns, only HSmax has slightly smaller values than the others so a clear distinction is difficult. Other than HNmax, HSmax is not solely dependent on one heavy precipitation event, but also on the already existing snow pack. This is true for HSavg as well, however, HSavg might incorporate seasonal differences within a station pair due to varying exposure to sun light towards the end of the snow season, whereas HSmax does not. In terms of RMSE, the HN-derived indicators score slightly better than the HS-ones except HNsum. It is no surprise to see such high values for the RMSE of HNsum as HNsum itself is an order of magnitude greater than the rest of the indicators. However, the HN-indicators are generally smaller and therefore the small RMSE can be misleading as it is sensitive to the scale of the actual values. In general, HN-indicators consist of fewer and most of the time smaller values than HS-indicators. In conjunction with a measurable resolution of 1 cm (ruler, average of three measurements, round up to nearest cm) measurement uncertainties attributed to small values (mainly HN) may have larger absolute effects.

4.3 | Days with snow cover and days with snowfall

As an indicator for snow-covered ground dHS1 is widely used. Would a threshold of 5 cm be more stable when considering possible breaks in the long-term measurements? The reason for 5 cm rather than 10, 20, or 50 cm is that even at 400 m a.s.l. values of HS and HN greater than 5 cm are frequently observed, whereas a threshold of 20 or 50 cm would only cover very few events which would dramatically reduce the sample size and inhibit any analysis.

The indicator dHS1 is most likely skewed and certainly capped at the total number of days between November and April for snow rich stations and years (very few occasions in our data set). However, this only applies to stations at high elevations where there is either enough snow all season or the temperatures are cold enough to prevent the snow from melting. dHS5 does not necessarily solve this problem because the threshold of 5 cm is still regularly exceeded at higher elevation stations. However, this 5 cm margin addresses another issue in connection with small values of snowfall or snow depth; the sensitivity of the exact measurement time. Given the right conditions, a snowfall of 1 cm can easily

melt or fall in 30 minutes and if the snowfall event takes place around the time of measurement it can be missed or recorded, solely dependent on the time the observer went out and conducted the measurements. Looking at the results, dHS1 and dHS5 are practically indistinct. Based on these results, there is no preference to either dHS1 or dHS5.

The same is true for dHN1 and dHN5; they both seem interchangeable when looking at COR, RPD, and RMSE. The only differences occur when looking at RC, where dHN5 (as well as dHS5) show stronger values than dHN1 (or dHS1) and RPD, where dHN1 shows a larger variability than dHN5, which is down to the temporal and spatial sensitivity of small precipitation events.

4.4 | Relative changes

RC only focus on a specific time window (1980–2004), which was chosen to maximise data availability. They serve mainly as additional information for inter-station-pair agreement and robustness. However, they generally agree with findings in Marty (2008) and Scherrer et al. (2013) of a decline in snow days in the Alps for the period after 1980. Within a station pair, the RC for various snow climate indicators are well in alignment with regards to the direction and strength, except for a few very weak RC that fluctuate around zero. Station pairs associated with large differences in RC for any indicator, visible as outliers in Figure 7, do not share any specific characteristics. For example, HSavg, which has the least similar RC per station pair, as reflected in the larger spread in Figure 7 and subsequently the largest MAD (Table 2). Low HSavg values normally imply less snow, which can particularly amplify the differences between stations at lower elevations where conditions can quickly change between freezing and melting, leading to short-lived snow appearances.

The stronger RC for dHS5 and dHN5 (compared to dHS1 and dHN1) highlight that a decrease in amount does not necessarily mean a decrease in frequency.

The overall smaller RC for HNmax can be explained by the large natural variability of the intensity of heavy precipitation events.

4.5 | Differences between Indicators

The memory effect of HS has an influence on persistence. Every HS at a specific day is to a certain amount dependent on the value of the day before and therefore more conservative, inertial, and stable in itself; whereas HN is a time series of (mostly) independent events. In contrast

to HSmax which is usually only dependent on accumulation, HSavg covers the entire season. Furthermore, HN-indicators such as HNmax, HN3max, dHN1 and dHN5 rather represent short and independent events that predominantly rely on precipitation which can vary on very small spatial scales; whereas HS-indicators are dependent on a combination of temperature and precipitation events therefore making them more inertial and apparently slightly more stable. However, HNsum and HN3max, which depend on a series of events as well showcase the most robust behaviour among the HN-indicators. Measurement uncertainties might have an impact as well, especially for small values (HN likely more affected than HS) as instruments and instructions do not permit for lower resolutions than 1 cm.

5 | CONCLUSIONS

When looking for the most robust indicators with regards to local changes, the following criteria have to be met: high COR, low RPD, small DRC, and low MAD for a majority of station pairs.

As the correlations for all indicators and a majority of station pairs is high (median COR > 0.8), only RPD, DRC and MAD are used to determine the most robust snow climate indicators. Low RPD are found in HSmax, dHS1, dHS5, and HNsum (closely followed by HN3max). Taking DRC into consideration has little effect as the values are similar for the snow climate indicators in question. The variability of the various RPD and DRC is quantifiable with the robust measure MAD (see Section 3.3 and Table 2). Smaller values indicate smaller variability and therefore imply a more robust indicator. Generally, DRC show larger MAD (Figure 3, 4, and 7 and Table 2). The most stable snow climate indicators with regards to local changes considering RPD, DRC, and MAD are: HSmax, dHS1, dHS5, and HNsum closely followed by HN3max. Surprisingly, HSavg, is according to our analyses, not among the most robust snow climate indicators and perhaps not the indicator to go to for future homogenisation efforts for snow, as it was used in Marcolini et al. (2017a).

Our analysis shows that median RMSE are about 5 cm for all height/depth related indicators and about 5 days for all time related indicators, except HNsum. Median relative percentage differences are about 7% for the number of days with snow cover and 11–16% for all other indicators. It is worth bearing in mind that in extreme cases the deviations within a station pair can reach 25–40%.

The differences between dHS1 and dHS5 respectively dHN1 and dHN5 are negligible except for the fact that both dHS5 and dHN5 show stronger changes than dHS1

and dHN1 and dHN1 has a larger variability than dHN5. A higher threshold of 5 cm (compared to 1 cm) does provide more stability on days with snowfall, but not on days with snow on the ground.

Nearly all stations show negative RC for all snow climate indicators during the period 1980 to 2004, suggesting that the climate signal during that period is stronger than the local inhomogeneities. But more importantly, the signs of strong RC agree at all station pairs for all the snow climate indicators.

All outliers in Figures 3–5 (correlation, relative differences and RMSE) can be attributed to the same five station pairs. However, the outliers in Figure 7 (difference in relative changes) show a more diverse pattern and consist of 10 different station pairs (including three from Figures 3–5).

ACKNOWLEDGEMENTS

We thank the two anonymous reviewers for their comments which helped to improve the manuscript.

ORCID

Moritz Buchmann  <https://orcid.org/0000-0002-1710-5496>

Michael Begert  <https://orcid.org/0000-0002-9569-6842>

REFERENCES

- Acquaotta, F., Colombo, N., Fratianni, S., Romeo, V. and Barbero, S. (2015) Preliminary results of a comparison study between two independent snow networks in north-western Italian Alps (Piemonte Region). In: Lollino, G., Manconi, A., Clague, J., Shan, W. and Chiarle, M. (Eds.) *Engineering Geology for Society and Territory - Volume 1*. Cham: Springer International Publishing, pp. 113–116.
- Acquaotta, F., Fratianni, S., Aguilar, E. and Fortin, G. (2019) Influence of instrumentation on long temperature time series. *Climatic Change*, 156(3), 385–404. <https://doi.org/10.1007/s10584-019-02545-z>.
- Acquaotta, F., Fratianni, S. and Venema, V. (2016) Assessment of parallel precipitation measurements networks in Piedmont, Italy: Assessment of parallel precipitation measurements networks in piedmont. *International Journal of Climatology*, 36 (12), 3963–3974. <https://doi.org/10.1002/joc.4606>.
- Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J. 2003. Guidance on metadata and homogenization, 54.
- Baronetti, A., Fratianni, S., Acquaotta, F. and Fortin, G. (2019) A quality control approach to better characterize the spatial distribution of snow depth over New Brunswick, Canada. *International Journal of Climatology*, 39(14), 5470–5485. <https://doi.org/10.1002/joc.6166>.
- Begert, M., Schlegel, T. and Kirchhofer, W. (2005) Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *International Journal of Climatology*, 25(1), 65–80. <https://doi.org/10.1002/joc.1118>.
- Beniston, M., Rebetez, M., Giorgi, F. and Marinucci, M.R. (1994) An analysis of regional climate change in Switzerland.

- Theoretical and Applied Climatology*, 49(3), 135–159. <https://doi.org/10.1007/BF00865530>.
- Della-Marta, P.M. and Wanner, H. (2006) A Method of Homogenizing the Extremes and Mean of Daily Temperature Measurements. *Journal of Climate*, 19(17), 4179–4197. <https://doi.org/10.1175/JCLI3855.1>.
- Gubler, S., Hunziker, S., Begert, M., Croci-Maspoli, M., Konzelmann, T., Brönnimann, S., Schwierz, C., Oria, C. and Rosas, G. (2017) The influence of station density on climate data homogenization. *International Journal of Climatology*, 37(13), 4670–4683. <https://doi.org/10.1002/joc.5114>.
- Guenzi, D., Acquaotta, F., Garzena, D., Baronetti, A. and Fratianni, S. (2020) An algorithm for daily temperature comparison: Co.Temp - comparing series of temperature. *Earth Science Informatics*, 13(1), 205–210. <https://doi.org/10.1007/s12145-019-00414-y>.
- Haberkorn A (ed). 2019. European Snow Booklet.
- Hunziker, S., Gubler, S., Calle, J., Moreno, I., Andrade, M., Velarde, F., Ticona, L., Carrasco, G., Castellón, Y., Oria, C., Croci-Maspoli, M., Konzelmann, T., Rohrer, M. and Brönnimann, S. (2017) Identifying, attributing, and overcoming common data quality issues of manned station observations: Identifying, attributing, and overcoming common data quality issues. *International Journal of Climatology*, 37(11), 4131–4145. <https://doi.org/10.1002/joc.5037>.
- Jonas, T., Rixen, C., Sturm, M. and Stoeckli, V. (2008) How alpine plant growth is linked to snow cover and climate variability. *Journal of Geophysical Research: Biogeosciences*, 113(G3), 2156–2202. <https://doi.org/10.1029/2007JG000680>.
- Kendall, M.G. (1975) *Rank Correlation Methods*. London: Charles, Griffin.
- López-Moreno, J.I., Revuelto, J., Fassnacht, S.R., Azorín-Molina, C., Vicente-Serrano, S.M., Morán-Tejeda, E. and Sextstone, G.A. (2015) Snowpack variability across various spatio-temporal resolutions: SNOWPACK VARIABILITY ACROSS VARIOUS SPATIO-TEMPORAL RESOLUTIONS. *Hydrological Processes*, 29(6), 1213–1224. <https://doi.org/10.1002/hyp.10245>.
- Mann, H.B. (1945) Nonparametric tests against trend. *Econometrika*, 13, 245–259.
- Marcolini, G., Bellin, A. and Chiogna, G. (2017a) Performance of the standard normal homogeneity test for the homogenization of mean seasonal snow depth time series: Performance of SNHT for snow depth time series. *International Journal of Climatology*, 37, 1267–1277. <https://doi.org/10.1002/joc.4977>.
- Marcolini, G., Bellin, A., Disse, M. and Chiogna, G. (2017b) Variability in snow depth time series in the Adige catchment. *Journal of Hydrology: Regional Studies*, 13, 240–254. <https://doi.org/10.1016/j.ejrh.2017.08.007>.
- Marty, C. (2008) Regime shift of snow days in Switzerland. *Geophysical Research Letters*, 35(L12501), 5. <https://doi.org/10.1029/2008GL033998>.
- Morán-Tejeda, E., López-Moreno, J.I. and Beniston, M. (2013) The changing roles of temperature and precipitation on snowpack variability in Switzerland as a function of altitude. *Geophysical Research Letters*, 40(10), 2131–2136. <https://doi.org/10.1002/grl.50463>.
- Neumann, N.N., Derksen, C., Smith, C. and Goodison, B. (2006) Characterizing local scale snow cover using point measurements during the winter season. *Atmosphere-Ocean*, 44(3), 257–269. <https://doi.org/10.3137/ao.440304>.
- Resano-Mayor, J., Korner-Nievergelt, F., Vignali, S., Horrenberger, N., Barras, A.G., Braunisch, V., Pernollet, C.A. and Arlettaz, R. (2019) Snow cover phenology is the main driver of foraging habitat selection for a high-alpine passerine during breeding: implications for species persistence in the face of climate change. *Biodiversity and Conservation*, 28(10), 2669–2685. <https://doi.org/10.1007/s10531-019-01786-9>.
- Scherrer, S.C., Ceppi, P., Croci-Maspoli, M. and Appenzeller, C. (2012) Snow-albedo feedback and Swiss spring temperature trends. *Theoretical and Applied Climatology*, 110(4), 509–516. <https://doi.org/10.1007/s00704-012-0712-0>.
- Scherrer, S.C., Wüthrich, C., Croci-Maspoli, M., Weingartner, R. and Appenzeller, C. (2013) Snow variability in the Swiss Alps 1864–2009. *International Journal of Climatology*, 33(15), 3162–3173. <https://doi.org/10.1002/joc.3653>.
- Schmucki, E., Marty, C., Fierz, C., Weingartner, R. and Lehning, M. (2017) Impact of climate change in Switzerland on socioeconomic snow indices. *Theoretical and Applied Climatology*, 127(3), 875–889. <https://doi.org/10.1007/s00704-015-1676-7>.
- Sen, P.K. (1968) Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63, 1379–1389.
- Spreitzhofer, G. (1999) Spatial, Temporal and Intensity Characteristics of Heavy Snowfall Events over Austria. *Theoretical and Applied Climatology*, 62(3–4), 209–219. <https://doi.org/10.1007/s007040050085>.
- Theil, H. (1950) A rank-invariant method of linear and polynomial regression analysis. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen Series A*, 53, 386–392.
- Wipf, S., Stoeckli, V. and Bebi, P. (2009) Winter climate change in alpine tundra: plant responses to changes in snow depth and snowmelt timing. *Climatic Change*, 94(1), 105–121. <https://doi.org/10.1007/s10584-009-9546-x>.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Buchmann M, Begert M, Brönnimann S, Marty C. Evaluating the robustness of snow climate indicators using a unique set of parallel snow measurement series. *Int J Climatol*. 2020;1–11. <https://doi.org/10.1002/joc.6863>